

**Contrast measures for predicting perceptual effects of daylight
in architectural renderings**

S Rockcastle SMArchS, **ML Amundadottir** MSc ETH and **M Andersen** PhD

Interdisciplinary Laboratory of Performance-Integrated Design, Ecole Polytechnique Federale
de Lausanne, Lausanne, Switzerland

Received 7 January 2016; Revised 4 March 2016; Accepted

Daylit architecture is perceived as a dynamic luminous composition, yet most existing performance metrics were designed to evaluate natural illumination for its ability to adequately illuminate a two-dimensional task surface and avoid glare-based discomfort. It may be argued that task-driven approaches based on surface illumination and glare ignore the likelihood that contrast can provide positive impacts on our visual perception of space. Advances in these metrics to accommodate climate-based sky conditions and occupant behavior have improved our ability to evaluate task illumination and glare, yet the same attention has not been paid to evaluating positive perceptual responses to daylight. Existing studies have attempted to link subjective ratings of composition to simple global contrast metrics without reaching consensus. More advanced metrics have been developed in computational graphics and vision fields, but have not been applied to studies in qualitative lighting research. This paper introduces the results from an online experiment where subject ratings of daylight composition are compared to quantitative contrast measures across a series of renderings. This paper will identify which measures correlate to subjects' ratings of visual interest, and introduces a modified contrast algorithm, which can be used as a novel prediction model for visual interest in daylit renderings.

Address for correspondence: Siobhan Rockcastle, Interdisciplinary Laboratory of
Performance-Integrated Design, Ecole Polytechnique Federale de Lausanne, Lausanne,
Switzerland

EPFL ENAC IA LIPID
LE 1 114
Station 18
CH 1015, Lausanne
E-mail: Siobhan.rockcastle@epfl.ch

1. Introduction

In spaces where daylight is a primary source of illumination, our visual perception of architecture is largely influenced by the dynamic state of surrounding environmental conditions. While the sun's position is intrinsically linked to latitude and time, sky type is sensitive to variable climatic and atmospheric conditions, both of which produce diverse compositions of light and shadow. In an essay on Steven Holl's architectural works, theoretician Stanford Kwinter writes, "*For Holl, architecture is the science of experience...Light is not itself the plenum of matter, but rather what reveals and conveys it, like water in the paper into which pigment is placed*"¹. While we understand the link between illumination and our ability to perceive spatial depth and material texture, the ephemerality of these effects under natural lighting conditions is far less intuitive and can produce un-anticipated and even surprising results. In his seminal book titled, *The Eyes of the Skin: Architecture and the Senses*, Juhani Pallasmaa states that "*In great architectural spaces, there is a constant, deep breathing of shadow and light; shadow inhales and illumination exhales light*"². Mary Ann Steane discusses the key differences between daylight and artificial sources in her book titled, *The Architecture of Light*. While artificial light sources can be carefully combined to match a desired composition, they produce a static effect and can never match the "*nuance of mood created by the time of day and the wonder of the seasons*"³.

The compositional effects of shadow, contrast, and light directionality are essential to the visual performance of architecture, and yet their effects are most often defined as qualitative. Research into measuring the impacts on human perception has been limited. Aside from the obvious challenges associated with measuring qualities that are inherently subjective, our current discourse on daylight performance has been dominated by energy-related concerns, brought on by the energy crisis of the 1970s and strengthened by the shift toward sustainable building practices. In an effort to reduce energy consumption, daylighting research has gravitated toward the widespread development of task-based illumination metrics to assess general illumination thresholds while minimizing electric energy use⁴. Visual comfort metrics, especially those pertaining to glare, have also gained influence in the last two decades, as the emphasis on daylight integration has led to an increase in glazed facades and complex shading systems that can trigger occupant discomfort in places where visual tasks are performed^{5,6}. While research in these areas is undeniably important, positive perceptual performance indicators took a back seat and only gained momentum in the last decade due to concerns that existing illumination-based metrics are not addressing light perceived from an occupant's field-of-view⁷.

In some ways, the idea of evaluating perceptual lighting quality through quantitative measures is somewhat superfluous. Why would we need to quantify the performance of something that we can readily evaluate using qualitative judgement? While people can observe and assess the visual effects of daylight in a single moment of time, they cannot intuitively comprehend or predict the range of effect that might be experienced over time. As daylight is a highly dynamic source, the complexity of

predicting performance necessitates a method that can evaluate a space over time and across diverse sun positions to communicate the variable impacts of light and shadow. Simulation is a powerful tool for evaluating performance dynamics as we can assess a range of temporally-induced effects. Existing tools are most commonly used to assess illumination and glare risk, and yet there are few if any dynamic simulation-based methods for evaluating the perceptual aspects of daylight composition or its potential effects on architectural design.

As will be discussed in Section 2.2, while there are studies linking global contrast metrics to perceived impressions of visual interest, more sophisticated local contrast measures exist in vision science and psychology. These, however, have not been used to evaluate the perceptual performance of daylit architecture. If we can use image processing to quantify contrast-based visual effects within a single rendering and successfully link these values to impressions of spatial composition and visual interest, then we can apply that measure to a series of hourly and daily instances and predict these effects over time. This would help designers to understand where (within a defined view) and when (across hourly and daily moments) the effects of contrast, light, and shadow are likely to produce specific responses.

In this paper, the authors will apply existing contrast metrics from vision science and psychology to high dynamic range (HDR) renderings for a series of nine contemporary architectural spaces under three different sunny sky conditions that were selected due to their variation in daylight composition. These metrics will then be compared to subjective ratings for contrast, uniformity, variation, direction, complexity, excitement, and stimulation that have been gathered through an online survey. The results from this experiment will then be used to derive a model for predicting visual interest using ordered logistic regression between subject ratings and a modified contrast algorithm.

2. Background

Studies that have assessed the perceptual impacts of contrast on daylit space have relied primarily on subjective surveys to explore the relationship between simple photometric measurements and perceived impressions of interior space⁸⁻¹¹. Two factors that are widely accepted to impact the field-of-view in daylit architecture are average luminance and luminance variation¹². The former has been directly associated with perceived brightness and the latter with visual interest¹³. Some studies have found that both mean luminance and luminance variation within an office environment contribute to occupant impressions of preference^{10, 11, 14}, whereas others have discovered that the distribution of luminance values across an occupant's field-of-view^{15, 16} as well as the strength of variation are factors of preference^{9, 17}.

2.1 Existing contrast measurements

The problem with studies that rely on simple photometric measures such as average luminance and luminance variation, is that they generally do not address the spatial diversity of luminance values within an occupants' field-of-view. In these studies, luminance variation or contrast is most commonly defined by a global measurement, such as Michelson or Root Mean Square (RMS) contrast.

The Michelson contrast¹⁸ is defined as

$$\text{Michelson} = \frac{p_{\max} - p_{\min}}{p_{\max} + p_{\min}}, \quad (1)$$

where pixel intensities p_{\max} and p_{\min} are two single points of extreme brightness and darkness in an image. The RMS contrast measures the root mean squared difference of individual pixel intensities from the mean¹⁹

$$\text{RMS} = \sqrt{\frac{1}{WH} \sum_{i=1}^W \sum_{j=1}^H (p_{i,j} - \bar{p})^2}, \quad (2)$$

where $p_{i,j}$ are the pixels intensities at position (i, j) in an image of size W by H and \bar{p} is the average pixel intensity.

These global contrast metrics provide a single comprehensible value, which existing studies in daylight perception have used for comparisons with subjective rankings⁹. Yet, they cannot effectively predict perceived contrast between two images that vary in composition²⁰. To overcome this limitation, more sophisticated contrast metrics have been developed in the fields of image analysis and vision research. The current state of the art in these fields defines two types of metrics that are commonly used to quantify contrast: those that rely on global metrics (such as Michelson and RMS) and those that rely on local metrics²⁰. Local contrast metrics were developed to overcome the limitations associated with global metrics by quantifying the effect of composition, i.e. contrasting areas of brightness and darkness. Included within this group of metrics are methods that evaluate spatial frequencies in the Fourier domain²¹, those that measure a weighted colour contrast based on the distance between chroma regions²², and those that calculate the difference between a single pixel and a surrounding region or neighbourhood²³⁻²⁶. The authors have focused on the latter group of neighbourhood metrics for their ability to quantify the local contrast values between pixels within a neighbourhood or sub-region and assign a singular measure which represents the strength of local variation across all pixels. Spatial Contrast (SC), a measurement adapted from reference 24 and simplified by the authors for ease of computing²⁶, measures the sum of local pixel variations across a single image.

$$\text{SC} = \frac{1}{WH} \sum_{i=1}^W \sum_{j=1}^H \overline{\Delta p}_{i,j}, \quad (3)$$

where $\overline{\Delta p}_{i,j}$ is the average difference between the four pixels p_k orthogonally surrounding the central pixel $p_{i,j}$ or

$$\begin{aligned}
\overline{\Delta p}_{i,j} &= \frac{1}{4} \sum_{k \in K_4} |p_{i,j} - p_k| \\
&= \frac{1}{4} (|p_{i,j} - p_{i+1,j}| + |p_{i,j} - p_{i-1,j}| + |p_{i,j} - p_{i,j+1}| + |p_{i,j} - p_{i,j-1}|)
\end{aligned} \tag{4}$$

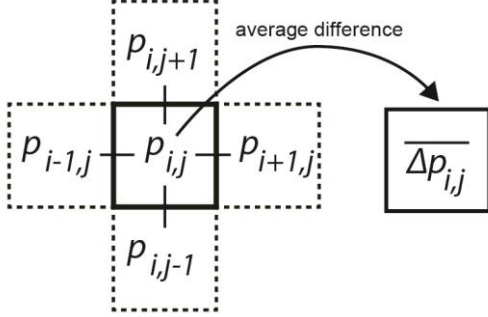


Figure 1 Pixel diagram showing $\overline{\Delta p}_{i,j}$, the average difference between each pixel $p_{i,j}$ and the four pixels orthogonally surrounding $p_{i,j}$, see Equation (4).

RAMMG, a contrast algorithm developed by Rizzi *et al*²⁴ and named after the authors who developed it, applies a multi-level approach to compute mean local pixel variations across a pyramidally subsampled structure, taking into account perceived differences in brightness across multiple image resolutions. The overall measure for N number of levels is described as

$$\text{RAMMG} = \frac{1}{N} \sum_{l=1}^N \bar{c}_l, \tag{5}$$

where \bar{c}_l is the mean contrast in the level l

$$\bar{c}_l = \frac{1}{WH} \sum_{i=1}^{W_l} \sum_{j=1}^{H_l} c_{i,j}. \tag{6}$$

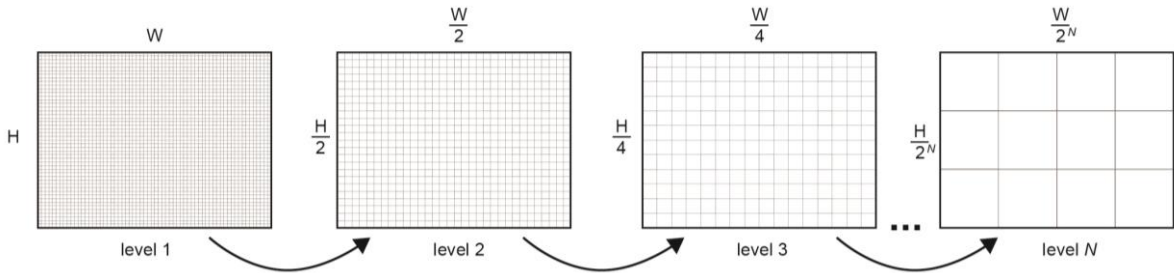


Figure 2. Diagram showing the reduction of image resolution $H \times W$ at each subsequent level l according to the multi-level approach RAMMG.

The image resolution is halved in each subsequent level, where $W_l = W_{l-1}/2$ and $H_l = H_{l-1}/2$ are the width and height of the image at level l and $c_{i,j}$ is the contrast of each pixel, calculated as

$$c_{i,j} = \sum_{k \in K_8} \alpha |p_{i,j} - p_k|, \tag{7}$$

where pixels p_k are the 8 neighbouring pixels of $p_{i,j}$ and the weight α applied to each of the 8 surrounding pixels k is

$$\alpha = \frac{1}{4+2\sqrt{2}} \begin{bmatrix} \frac{\sqrt{2}}{2} & 1 & \frac{\sqrt{2}}{2} \\ 1 & & 1 \\ \frac{\sqrt{2}}{2} & 1 & \frac{\sqrt{2}}{2} \end{bmatrix}. \quad (8)$$

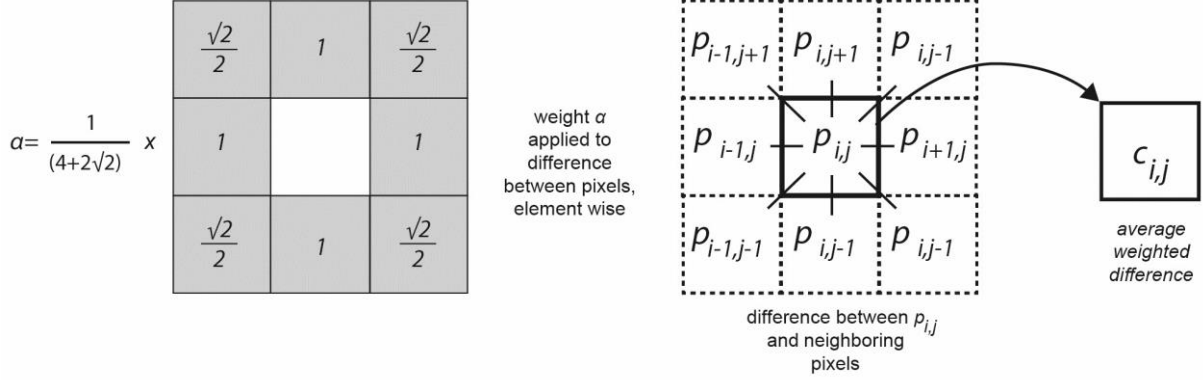


Figure 3. Diagram showing c_{ij} , the average difference between $p_{i,j}$ and the 8 neighbouring pixels, multiplied by a weight α , see Equation (7).

Multi-level metrics like RAMMG were developed under the assumption that a multiresolution analysis can better explain different mechanisms of perception in the human visual system^{27,28}. Whereas large image resolutions (>100,000 pixels) provide the detail to compute small, localized contrast evaluated between pixel neighbours, small image resolutions (<25,000) provide the opportunity to measure the difference between larger areas of brightness (i.e. larger neighbourhoods). While a pixel does not directly translate to the solid view angle between a subject and 3-dimensional scene due to the flattening of pixels in a 2D image, larger pixels may be seen as representing a larger view angle to the scene. Large image resolutions may therefore correspond to higher-level vision functions such as object and pattern recognition, while small image resolutions may correspond to lower-level functions such as the extraction of light stimulation to construct surfaces, edges, and features. The authors believe that some resolution levels may relate to contrast-driven subjective responses more than others, which is also suggested by Rizzi *et al.* based on their findings²⁹ as are discussed in Section 4.3.

The Difference of Gaussian (DOG) metric²³ differs substantially from the previous three contrast metrics in that it computes local differences between two bi-dimensional Gaussian filters with a centre component $R_c(x, y)$ and a surround component $R_s(x, y)$

$$\text{DOG}(x, y) = \frac{R_c(x, y) - R_s(x, y)}{R_c(x, y) + R_s(x, y)}. \quad (9)$$

The computation of centre and surround components $R_c(x, y)$ and $R_s(x, y)$ can be found in the Appendix. In the original metric, the final calculation was computed as an average of $\text{DOG}(x, y)$ taken

across 1,000 randomly selected pixels within an image to save computational time. The authors have computed the average $DOG(x, y)$ across all pixels in a given image with width W and height H

$$\overline{DOG} = \sum_{i=1}^W \sum_{j=1}^H DOG(x_i, y_j). \quad (10)$$

In 2009, Simone *et al.* combined the multi-level approach developed for RAMMG and the DOG metric to create a multi-level metric called RSC²⁸

$$RSC = \frac{1}{N} \sum_{l=1}^N \overline{DOG}_l, \quad (11)$$

where N is the number of levels and \overline{DOG}_l is the mean contrast in each level l

$$\overline{DOG}_l = \sum_{i=1}^{W_l} \sum_{j=1}^{H_l} DOG(x_i, y_j). \quad (12)$$

Each of these metrics will be applied to a selection of 27 images to compare their response to subjective ratings of daylight composition, gathered through an online survey conducted by the authors.

2.2 Existing experimental studies

To conduct qualitative lighting research using digital images, existing studies have applied subjective rating methods to measure impressions of lighting composition in HDR photographs^{11,30-32} as well as rendered images of a simulated office environment¹⁰. These experiments have asked participants to view a series of images and then respond to semantic differential ratings⁸ for pleasantness, contrast, brightness, spaciousness, and/or distribution, which were then compared to photometric measurements taken from the digital images.

When using images to collect subjective impressions of daylight related to perceptual factors such as brightness and contrast, it is essential that light levels are accurately captured or generated (in the case of renderings) and displayed with as broad a range of luminance levels as possible using proper tone-mapping algorithms calibrated for the specific display. In controlled laboratory experiments, tone-mapped HDR images have been displayed to subjects using 2D or 3D projection, HDR displays, and conventional low dynamic range (LDR) displays. There are now backlit HDR screens on the market which can display luminance values up to 4,000 cd/m^2 ³³, but a study by Cauwerts in 2013 found no significant difference in subjective assessments of contrast between images displayed on HDR displays and those that were tone-mapped to conventional LDR displays. In this study there were some differences in ratings of pleasantness, distribution, and spaciousness between real world scenes and image displays, however it was concluded that conventional LDR displays of 200 cd/m^2 (with images tone mapped to 256 distinct luminance levels) could be used as a surrogate for real world spaces to conduct subjective assessments involving contrast and brightness³⁴. In 2012, Villa and Labayrade developed an online protocol to limit the impacts of uncontrolled experimental

conditions (i.e. screen resolution, brightness, background, etc.) on the assessment of digital images for lighting quality research. Their study found that significant effects could be identified with 40 subjects despite systematic error due to uncontrolled conditions and the variance in p-value was constant with ≥ 100 subjects (0.0002)³⁵. Where controlled experiments are time-consuming, an online protocol allows for more widely distributed experiments.

In summary, there are a number of methods for creating and displaying interior images to assess subjective qualities of daylight, each of which has its own set of advantages and limitations. While real spaces produce the most accurate impression of light for subjective assessment, experimental conditions are limited by the physical sky conditions available on site at the time and it is difficult to compare a range of spatial configurations or temporal conditions in an efficient manner. HDR renderings using Radiance³⁶ allow for a broad range of spatial and temporal lighting conditions, but are limited by the luminance output of the display device and must therefore apply tone-mapping algorithms to achieve an acceptable range.

3. Method

The authors designed and distributed an online survey to a large pool of subjects from around the world. This experiment was designed to test the effects of architectural composition and sun position on subjective responses for contrast, distribution, and excitement ratings and allow the authors to compare existing global and local contrast measures to these ratings. The objectives of this experiment are two-fold: 1) to estimate the impact of sky conditions and architectural spaces on subjective ratings of contrast-related characteristics in daylight composition, and 2) to compare the relationship between these subjective ratings and existing quantitative measurements for contrast. The first objective is to test whether the subjects reach a consensus on their ratings of contrast-based visual effects in architectural spaces and whether these ratings are sensitive to sunlight dynamics (sky types). The second objective is to compare the existing contrast measurements presented in Section 2.1 to subjective ratings to determine whether these quantitative metrics can be used as a model for predicting perceptual responses to daylight composition.

3.1 Architectural spaces

To generate quantitative data and gather subjective data, the authors modeled nine contemporary architectural spaces that display a range of contrast-based visual effects and were selected based on a prominent interior view. To select the architectural spaces used in this initial study, the authors evaluated a wide range of interior scenes and chose those examples which represent a unique typology of daylight composition within a broader gradient of contrast-based effects³⁷. As can be observed in Figure 4, we grouped spaces into three contrast categories. For the high contrast group, the authors selected the Arab World Institute by Jean Nouvel (*arab*), the Zolleverein School by SANAA (*zoll*), and the Serpentine Pavilion by Toyo Ito (*serp*). The medium group contains the

Neughebauer House by Richard Meier (*neug*), the Toldeo Glass Museum by SANAA (*toledo*), and the First Unitarian Church by Louis Kahn (*first*), all of which show a mix of direct and indirect daylight conditions. Finally, the low contrast group holds the Poli House by Pezo Von Ellrichshausen (*poli*), the Thermal Baths at Vals by Peter Zumthor (*vals*), and the Menil Gallery by Renzo Piano (*menil*). This group is characterized by more indirect daylight design.

Each of the selected spaces was modelled in Rhinoceros version 5 sr6 based on available building plan and section drawings, and exported to Radiance using the Diva 2.0 toolbar to produce HDR daylight renderings. The authors chose not to model furniture or temporary artifacts within each space in an attempt to limit visual obstructions and minimize biases toward programmatic function. To compress HDR images down to an acceptable range for conventional computer screens (0.5 to 200 cd/m²) i.e. typical for personal tablet, laptop and desktop screens, the PCOND mapping algorithm was used³⁸. While tone-mapping produces a range of pixel values acceptable for conventional screen displays, the authors acknowledge the potential limitations associated with a compressed range of values. As Cauwerts discusses in her experimental findings using HDR, LDR, and projection, a compressed range of luminance values can produce an acceptable surrogate for real spaces when testing subjective assessments of contrast and brightness³⁴. While the range of luminance values experienced in a real-world scene far exceed those that may be reproduced by an LDR or even HDR screen, the authors felt the advantage of wide-spread survey diffusion by means of conventional computer displays outweighs the marginal effect of display luminance range. To assess this assumption further, an expanded luminance range will be explored in a forthcoming laboratory-based experiment.

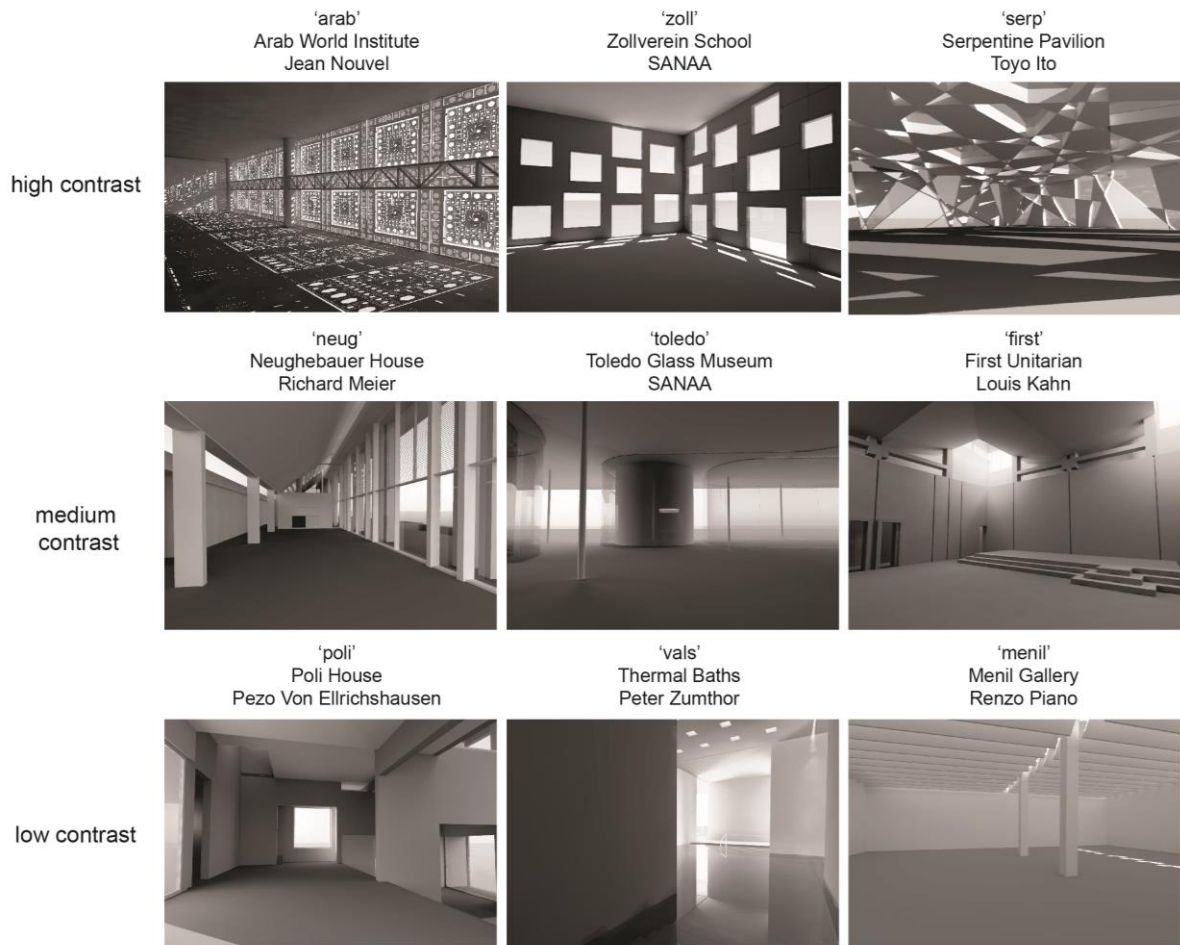


Figure 4. The nine architectural spaces selected for this experiment, grouped into high contrast, medium contrast, and low contrast categories.

3.2 Experimental Design

In this experiment, the authors have chosen a repetitive 3 x 3 Semi-Latin Square which allows for the comparison of three variables – spaces, subject group, and sky - while limiting experimental fatigue by showing each subject 9 images, rather than the 27 that are required by a full factorial experiment. The Semi-Latin Square allows for repetition (in the case of multiple variables within a given group) and nesting (with three architectural examples per sub category of high, medium, and low contrast – nine spaces in total). While the experiment is intended to study the effects of ‘space’ and ‘sky type’ on a subject’s impression of contrast-based factors within a daylight rendering, there is a risk that subjective assessments of daylight composition repeated under multiple sky conditions will create a bias toward the architecture. To overcome this, a Latin Square experimental design uses three distinct subject groups. Each subject within a group is shown a single rendering of each of the nine spaces, under one of the three possible sun positions. With three distinct groups, we can then compare the effect of architectural composition within groups and the effect of sun position between groups.

Figure 5 shows this logic. This proposed experimental design will increase response rate (due to faster completion times) and allow us to test the effect of sky type between groups.

To select the dates and times for each rendering within the study, the authors divided half the year (from the winter to summer solstice) into 28 moments, which represent symmetrical daily and monthly instances. Each of the nine architectural spaces was then rendered at each of the 28 moments and analyzed in MATLAB R2012b using the RAMMG contrast metric²⁴, which was selected to represent the broader group of neighbourhood metrics introduced in Section 2.1. From the assessment of RAMMG contrast across these 28 renderings, three images were selected: the highest (Sky 1), lowest (Sky 3), and mean (Sky 2) contrast composition for each space. The date and time for Sky 1, Sky 2, and Sky 3 therefore vary for each space as the RAMMG contrast assessment selected the highest, lowest, and mean compositions relative to the 28 renderings produced for each space. Based on the mean RAMMG contrast for each architectural space, the 9 spaces were ordered and divided into three sub-groups: high, medium, and low.

Both global (Michelson and RMS) and local contrast metrics (SC, RAMMG, DOG and RSC) were applied to the renderings in this study. DOG measurements are dependent on the radius of centre and surround Gaussian filters with a range of possible combinations. To limit this range, the authors applied a series of radii combinations from $r_c = 1$ to 4 and $r_s = 2$ to 8, based on combinations chosen in other experiments^{23,29}.

As metrics like RAMMG and RSC are applied to a series of levels depending on the original input resolution of the image, the authors also looked at each level independently. In this study, the original images were 1488 x 1024 pixels. We halved the resolution at each subsequent level, looking at 9 independent image levels for RAMMG (RAMM1, RAMM2, ... , RAMM9), and 5-6 levels for RSC, depending on the centre and surround radii.

From an initial comparison of these metrics, it is already apparent that the local contrast metrics such as Michelson produce less differentiated predictions of contrast between spaces and sky conditions, despite the highly varied composition of pixels in each image. To find which of the metrics described above can be used to best predict perceptual visual effects, the contrast results will be applied to each image and then compared to semantic ratings collected from the online experiment.

To test the 3 x 3 Latin-Square design, the authors conducted a small pilot study composed of 5 female and 4 male researchers in building performance, with varying competencies in architectural design, computational, and civil engineering. Each subject was randomly assigned a survey (from three possible groups) and asked to rate each rendering (shown in Figure 4) for six semantic differential ratings. Using this limited sample size of three subjects in each survey group, the authors were able to see a significant effect of space and sky type, with an insignificant effect of subject, suggesting that the experimental design was effective³⁹.

3 x 3 Latin square assignment of images to subject groups

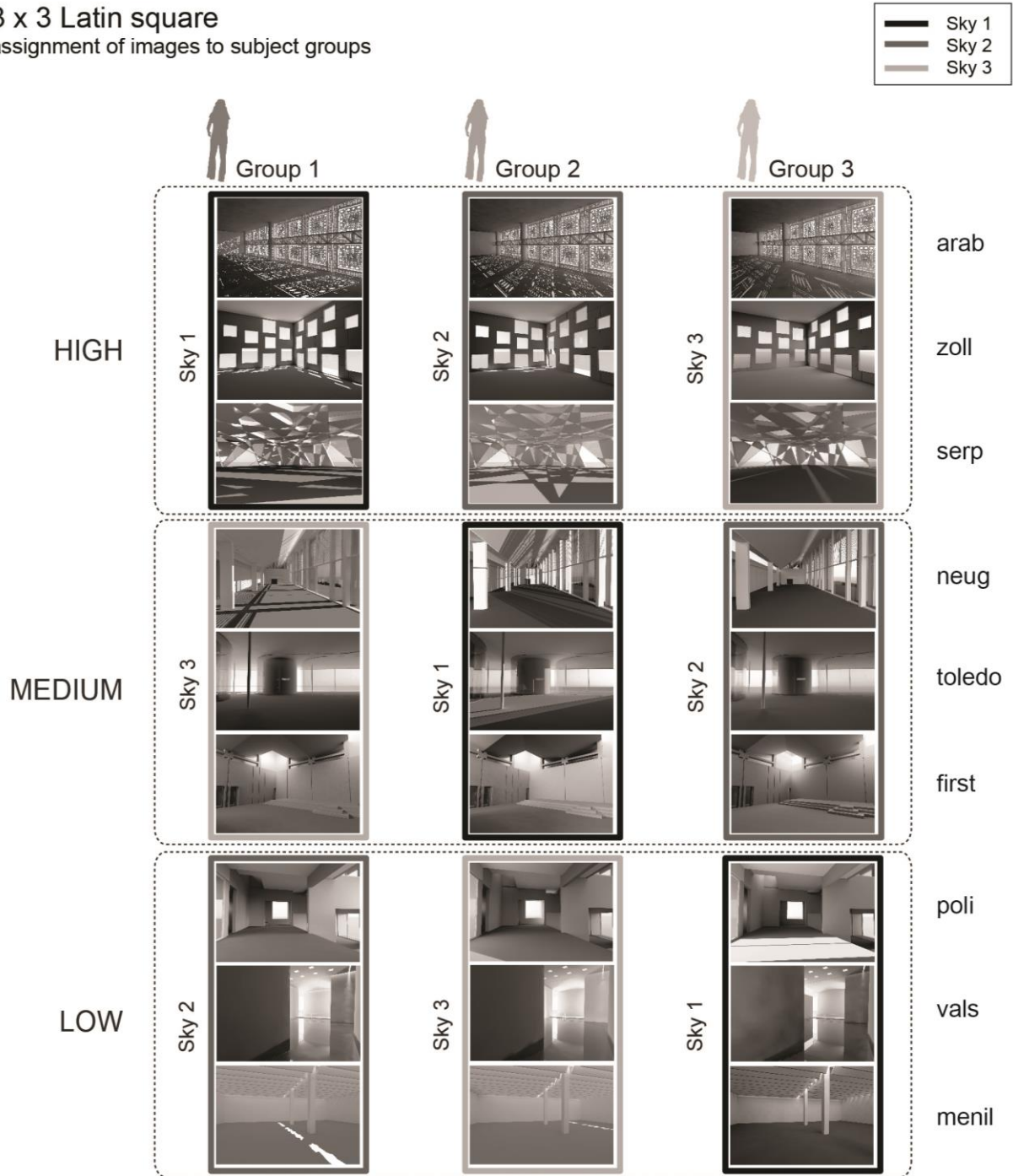


Figure 5. The renderings to each of the subject groups based on the 3 x 3 Latin Square design of experiment. Each group sees all nine architectural spaces under at least one of the three contrast conditions (high, mean, and low) selected by the RAMMG metric.

3.3 Simulation parameters

Due to the computational intensity of Radiance, the 28 renderings used to select highest, lowest, and mean contrast compositions were first rendered using low-quality ray-tracing and image

parameters. After each set of 28 low-quality renderings was complete, RAMMG was used to select three images for each space (high, mean, and low). These images were then re-rendered using high-quality parameters (-ps 2 -pt .05 -pj 0 -dj 0 -ds .15 -dt .05 -dc .75 -dr 3 -dp 512 -st .15 -ab 5 -aa .1 -ar 512 -ad 4096 -as 2048 -lr 8 -lw .005 -vv 60 -vh 80) with increased pixel resolution (1488 x 1024). As a result, there is some variation in RAMMG values between the low quality images, which were used to select moments, and the high quality renderings, which were shown to subjects and analyzed with each of the available contrast metrics. Not surprisingly, the increase in computational intensity had an effect on RAMMG in both indirect and direct lighting conditions. This discrepancy does not, however, affect the comparison between quantitative calculations and subjective ratings, as the low-quality images were only used to select annual moments and only high quality images were shown to subjects at random.

3.4 Experimental procedure

The online survey designed for this experiment was created using Survey Gizmo with a branch logic which allowed for random group assignment upon initiation of the survey. Each group was asked to respond to some basic demographic questions regarding geographic location and profession and then shown the nine architectural spaces at random, under one of the three possible sky conditions. For example, Group 1 (Figure 6) was shown three high contrast spaces under Sky 1, three medium contrast spaces under Sky 3, and the three low contrast spaces under Sky 2. While we asked the participants not to initiate the survey on a smartphone, we allowed tablet, laptop, and desktop computers. Participants were asked to turn the brightness on their device to maximum, to ensure that the maximum possible pixel range was observed.

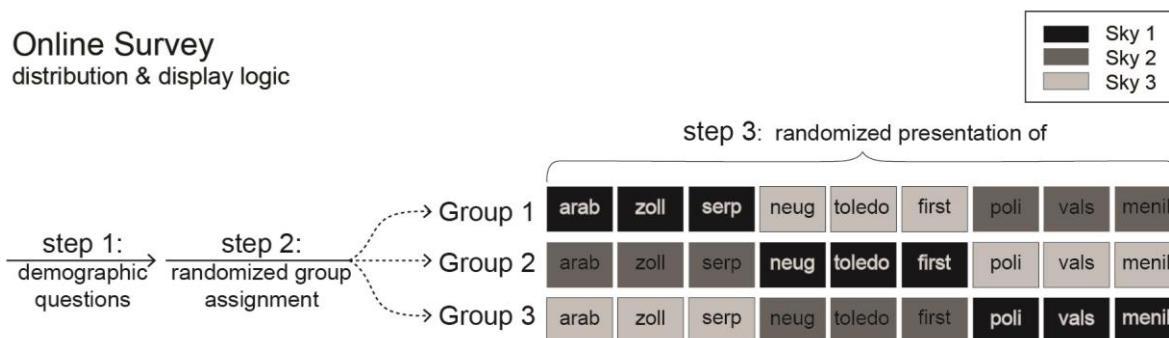


Figure 6. The online survey logic, where subjects are first introduced to basic demographic questions, after which they are randomly sorted into one of the three subject groups and asked to rate the images as they are presented in random order.

For each image, subjects were asked to rate the daylight composition using the following seven point semantic differential scales: low contrast – high contrast, uniform – non-uniform, unvaried – varied, diffuse – direct, simple – complex, calming - exciting, sedating – stimulating. Flynn

introduced the use of semantic differential scales to gather subjective assessments of daylight quality in terms of visual clarity, spaciousness, evaluation, relaxation, social prominence, complexity, modifying influence, and spatial modifiers⁸. Numerous studies thereafter have employed the use of these scales to conduct daylight quality research in real spaces and simulated or photographed views^{10,30,40,41}. For the proposed study, the authors have focused on bi-polar semantic differential scales associated with complexity and spatial modifiers as well as visual interest. The selection of a seven-point scale was motivated by the use of bi-polar pairs to allow for a neutral rating at 4 and three ratings in either direction of the scale. In our analysis, we were able to cluster subject ratings 1-3 and ratings 5-7 as being non-neutral and visualize the distribution of ratings on one side of the bi-polar scale or the other (as shown in Figure 7). Due to the limitations of Survey Gizmo in recording ordinal subject responses through the online survey format, bubble ratings were used instead of line scales. Pair-wise comparison was also considered, but rejected due to the number of rating pairs selected in our survey (seven pairs per image).

3.5 Data management

In total, there were 334 subjects who initiated the survey with 200 complete responses and 134 partially complete ones, which were discarded before the analysis. Interestingly, we did see a significant effect on responses from those subjects using tablets. These subjects (4.5%) were discarded as the authors believe that this effect could be due to the smaller screen size (which forced subjects to manually zoom in and out to fully view each image) or the default button format which was automatically adjusted by Survey Gizmo on the tablet version, switching from a 7-point rating scale with selection bubbles to a sliding scale. There was no significant effect observed between subjects using a laptop or a desktop computer. Of the remaining 175 subjects, 96% selected their English language capacity as professional, bilingual, or native, with the remaining 4% responding with elementary or limited working proficiency. These subjects were also discarded. From the remaining 168 subjects, 64% were designers (architecture, landscape, urban, or interior), 36 % non-designers, with 55% reporting their expertise in lighting design as competent, proficient, or expert, and the remaining 45% claiming novice or beginner expertise. One subject was excluded from the analysis because 73% of responses were neutral. We normalized the responses (from 1 to 7) for five other subjects, as they did not use either extreme on the rating scale. The remaining 167 subjects were evenly distributed among the three groups (Group 1: 55, Group 2: 56, Group 3: 56).

3.6 Data analysis

To test the significance of experimental factors on the data from each rating pair collected in the experiment, a 3-way analysis of variance (ANOVA) was used to test the effects of sky, space, and subject group. As the residuals for each rating pair was not normally distributed, a *post-hoc* analysis was conducted using the nonparametric Kruskal-Wallis test to determine the significance of each

group within the factor under consideration. To analyze the relationship between subject ratings and existing contrast metrics, the authors calculated Pearson's correlation coefficient. Using this coefficient, the authors selected those combinations of rating-pair and contrast measurement with $\rho \geq 0.69$ ($p < 0.0001$). An ordered logit model was then applied to fit the subjects' ratings to selected contrast measures, as the ratings are ordinal response scales. For a given category c on the semantic differential scale (1-7), the probability of each outcome is expressed as a nonlinear function of the predictor x

$$\ln \theta_j = a_j - bx, \forall j = 1, \dots, c. \quad (13)$$

where b is the corresponding effect parameter, a the cutoff point and θ_j the ratio of the probability that a response belongs to a category with a value less or equal to category j and the probability that this response belongs to a category with a value greater than category j .

$$\theta_j = \frac{P(y \leq j)}{P(y > j)}. \quad (14)$$

The cumulative predictive probabilities from the logit model for each rating are calculated as

$$P(y \leq j) = \frac{1}{1 + e^{a_j - bx}}. \quad (15)$$

4. Results

The description of results will be introduced in three parts. First, we will explore the distribution of subject responses for each semantic rating pair collected during the online experiment. Second, we will test the significance of experimental factors on each rating pair to determine whether the effects were statistically significant. Third, we will compare the relation between subject ratings and quantitative contrast measures and introduce a model for predicting visual interest in dynamic daylight architecture.

4.1 Distribution of subject responses

Figure 7 shows stacked bar plots with the distribution of subject responses for each level of the seven-point rating scale under each image presented in the experiment (organized by architectural space and sky type). Subject responses are sorted and clustered into gradients by tone gradient, with light grey demarcating ratings that fall on the left side of the scale (1-3) and dark grey for ratings that fall on the right (5-7). Ratings in the middle (4), indicating neutrality or indifference toward either side of the semantic scale are shown as white void. For example, ratings 1, 2 and 3 on the contrast spectrum are shown in a gradient of light grey, with 1 being the lowest contrast rating. A rating of 4 was perceived as neither low nor high in contrast and shown as white void. Ratings 5, 6 and 7 are shown in dark grey, with 7 as the highest possible contrast rating. The most frequent responses are

displayed for each rating as the percentage of subjects on the respective side of the rating scale. The dashed line shows where the median falls on the scale within each question. Using this graphic representation, we can see the relative distribution of subject responses for all ratings. Images with dominant bar length of dark grey or light grey (over 50%) can be considered partial toward one side of the rating scale. Images with large areas of void can be considered neutral between each end of the rating scale, while images with an even distribution of light and dark grey lacked consensus between subjects.

There is an obvious effect of sky type in some, but not all spaces – specifically those that see the largest variation in daylight composition due to sky type. *Zoll*, *neug*, and *vals* experience a noticeable shift in ratings between sky types 1, 2 and 3. The spaces with strongest subject consensus toward the light grey end of the rating scale (low contrast, uniform, unvaried, diffuse, simple, calming, subdued) are *first* and *menil*, while the dark grey side of the rating scale (high contrast, non-uniform, varied, direct, complex, exciting, stimulating) was dominated by *serp* and *arab*. While all rating scales were found to be significantly correlated to each other, the percentage of subjects that responded with ratings 5-7 for excitement and stimulation were the most highly correlated ($\rho = 0.98$).

4.2 Effects of experiment

The significance of experimental factors was evaluated using a 3-way ANOVA to test the effects of subject group, space, and sky type on each rating scale. While the ANOVA revealed a significant effect of both space and sky factors for all rating scales ($p < 0.01$), the residuals were not normally distributed. A *post-hoc* analysis was conducted using the non-parametric Kruskal-Wallis test, to study pair-wise comparisons between levels in each factor under consideration within each rating. This test was run for both sky type and space group on each of the semantic rating pairs. This test revealed the effect of sky was significant on subject responses to all rating scales ($p < 0.01$), except unvaried-varied. Figure 8 illustrates a pair-wise comparison between Sky 1 and Sky 3 which shows a significant effect ($p < 0.01$) on ratings of contrast, uniformity, direct, complexity, excitement, and stimulation. Ratings of excitement and stimulation also show a significant effect ($p < 0.01$) between Sky 1 and Sky 2, which suggest that these ratings were more sensitive to the range of sky types presented in this experiment. To test the effect of space, we grouped the examples into high, medium, and low based on the ordering in Figure 7. In this test, there was a significant effect of space between all groups in the factor ($p < 0.001$) for all rating pairs.

Stacked Bar Plot
Distribution of subject responses for each rating pair



Figure 7. Stacked bar plot showing distribution of semantic differential scores for each pair-wise rating and image. The images are ordered row-wise for sky type and column-wise based on the average percentages in categories 5-7 for all ratings and sky types per space.

Kruskal-Wallis pair-wise comparisons

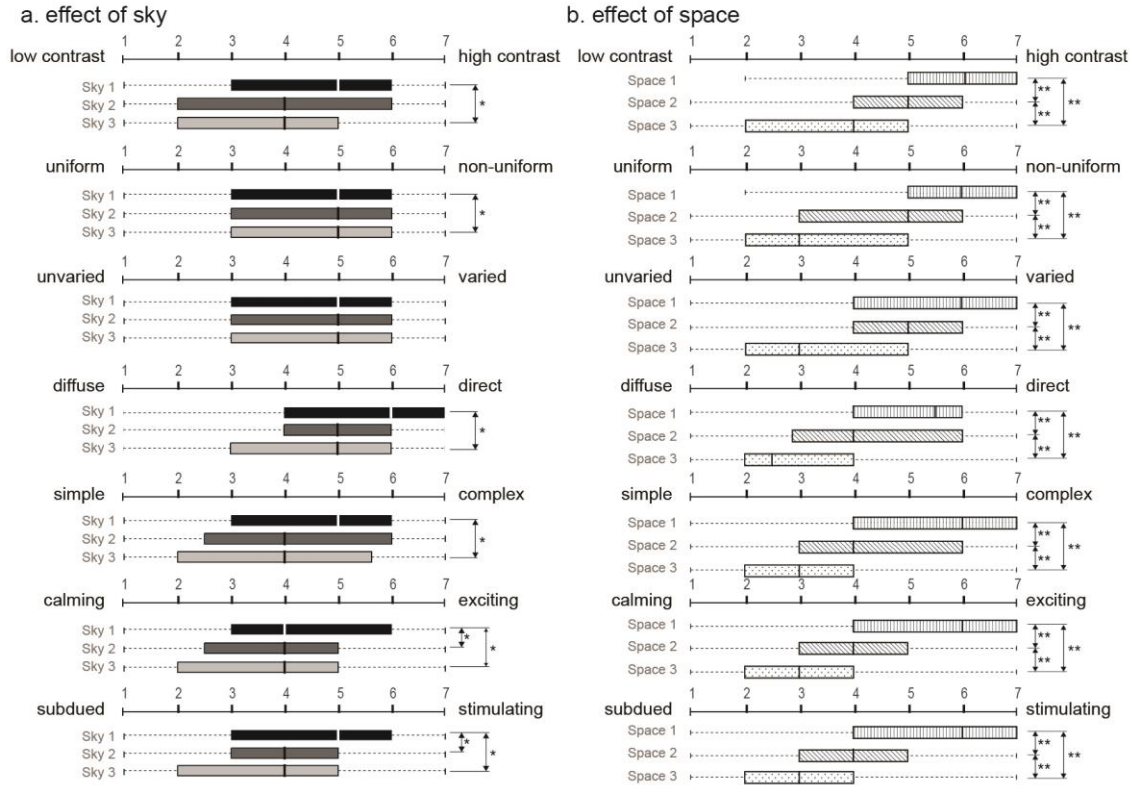


Figure 8. Kruskal-wallis pair-wise comparisons for the effect (a) between sky types 1,2 and 3 and (b) between space groups 1,2 and 3 within each rating scale.

This *post-hoc* analysis, in addition to the 3-way ANOVA would indicate that both space and sky factors had a significant effect on subject responses for all rating scales, except for variation. While the effect of space was significant on subject responses for all rating scales, the effect of sky was most significant on subject ratings of excitement and stimulation.

4.3 Subject ratings versus quantitative measures

To relate subject responses for each rating pair to the contrast measures described in Section 2.2, quantitative contrast results were compared to subject responses for each rating and image presented in the experiment. Table 1 (see Appendix) shows the correlations between the percentage of subjects that responded with ratings 5-7 for each image on each rating scale and the output for global and local contrast metrics. The Pearson correlation coefficient is shown for each rating/metric combination, with highlighted combinations for significance $p < 0.0001$.

In this experiment, ratings of contrast were found to have the strongest dependence with RAMMG ($\rho = 0.72$), but the RAMM of one image resolution level (RAMM5 - 64 x 93 pixels) in particular had the highest predictive fit to several other rating pairs. RAMM5 achieved the strongest

linear dependence to ratings of direct ($\rho = 0.80$), complex ($\rho = 0.72$), exciting ($\rho = 0.79$), and stimulating ($\rho = 0.78$). This finding suggests that the luminance resolution or radius of solid viewing angle may be related to subject impressions and tied to mid-levels of the visual system related to feature inference). Using Pearson correlation to pre-select contrast metrics as possible predictors of visual interest, we selected RAMM5, hereafter referred to as ‘Modified Spatial Contrast’ to study in more depth.

4.4 Towards an objective predictor for visual interest in architecture

As the subjects’ responses in this experiment were gathered using an ordinal bi-polar rating scale (1-7), the authors selected a probabilistic rather than definitive or threshold-driven approach. Existing illumination metrics rely on target or threshold illuminance to describe light levels that promote optimal visual acuity for a given task. In daylight glare probability (DGP), discomfort is calculated as a percentage of people that would rate a given glare source as either disturbing or intolerable. This model was established by grouping equal sample sizes into ‘classes’ from the total number of cases and evaluating the percentage of disturbed subjects in each class⁶. Where illumination is relatively objective, glare prediction must account for a range of subject responses to any given light condition. Unlike glare prediction, however, that relies on a uni-polar scale (from imperceptible to intolerable), the rating pairs used in this experiment are bi-polar (i.e. calming – exciting). That is to say that, unlike glare, neither side of the any rating pair is worse than the other.

Existing studies that have linked global contrast measures such as RMS to preference factors have found weak linear correlations and have not been robust enough to propose a predictive model for visual interest. The modified spatial contrast metric (RAMM5) and data from this experiment suggest that such a predictive model could exist and could indeed be robust enough to differentiate between architectural spaces that vary in daylight composition. An ordered logit model has been applied to fit the subjective ratings to RAMM5 (predictor x) using ordered logistic regression. Figure 9a shows the distribution of subject ratings ordered by RAMM5 and application of a predictive proportional odds model to subject ratings of calming – exciting in Figure 9b. The plot was obtained by using Equation (15) and the fitted parameters $a = [-0.45; 0.84; 1.67; 2.82; 3.64; 4.81]$ and $b = 0.24$. To evaluate the goodness of fit for this model, we looked at the significance of each cumulative response (1-7). At level 3, the odds of achieving ratings 1-3 are significant at $p < 0.05$.

When we group semantic differential ratings into two categories of calming (1-3) and exciting (5-7) as shown in Figure 9a, Figure 9b shows that a Modified Spatial Contrast of 13 (or more) triggers responses of excitement (ratings of 5, 6, or 7) for 63% of subjects, whereas a Modified Spatial Contrast of 5 (or less) produces responses of calming (ratings of 1, 2, or 3) in 59% of subjects. While we cannot and should not define one threshold or target value to predict perceptual impressions of

excitement given the subjective nature of such responses, this probabilistic model provides a first objective predictor for visual interest in daylight architecture.

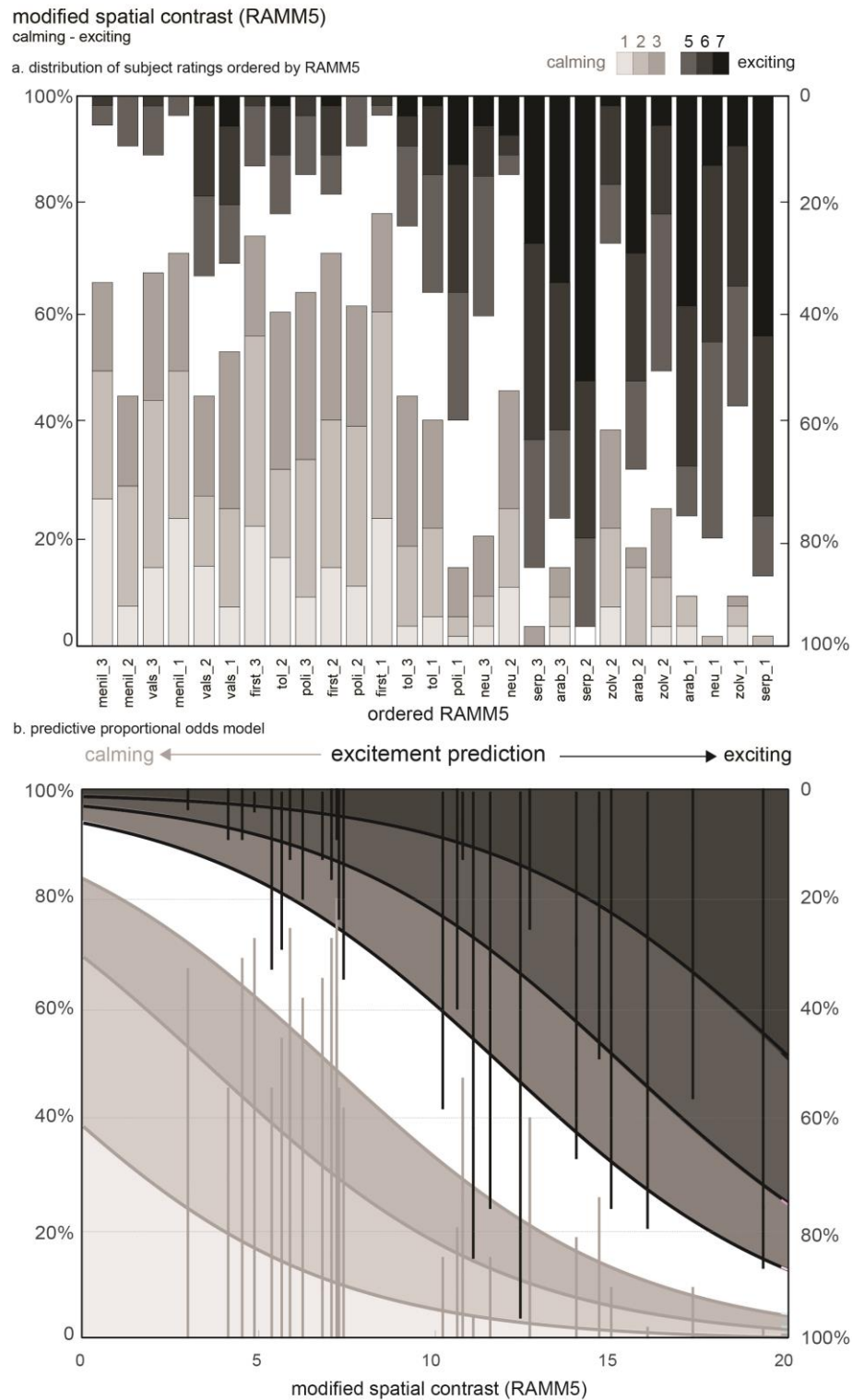


Figure 9. a) Distribution of subjects' ratings for calming – exciting per image ordered by RAMM5 and (b) predictive proportional odds model where the bars show subject responses (binned into light grey for ratings 1-3 and dark grey for ratings 5-7) for the 27 images located on the x axis at the calculated RAMM5. The curved lines show the predictive proportional odds through categories 1-6, where category 7 reaches 100%.

Figure 10 shows the application of modified spatial contrast to each of the spaces included in the experiment, rendered across 56 annual instances and mapped temporally to show an annual prediction for impressions of calming and exciting. In this figure, results are ordered by the average modified spatial contrast across all 56 annual instances with values in dark grey showing point-in-time predictions of excitement and light grey showing predictions of calming. As the experimental results have shown, there are modified spatial contrast values which can be identified as neutral, triggering consensus in neither impressions of calming or exciting, but for spaces where a large degree of variability occurs, this annual prediction can illustrate specific times of day and year when subjects are likely to identify the space on one side of the spectrum or the other. The darker grey to black values represent a stronger probability that exciting daylight compositions will be observed at that moment.

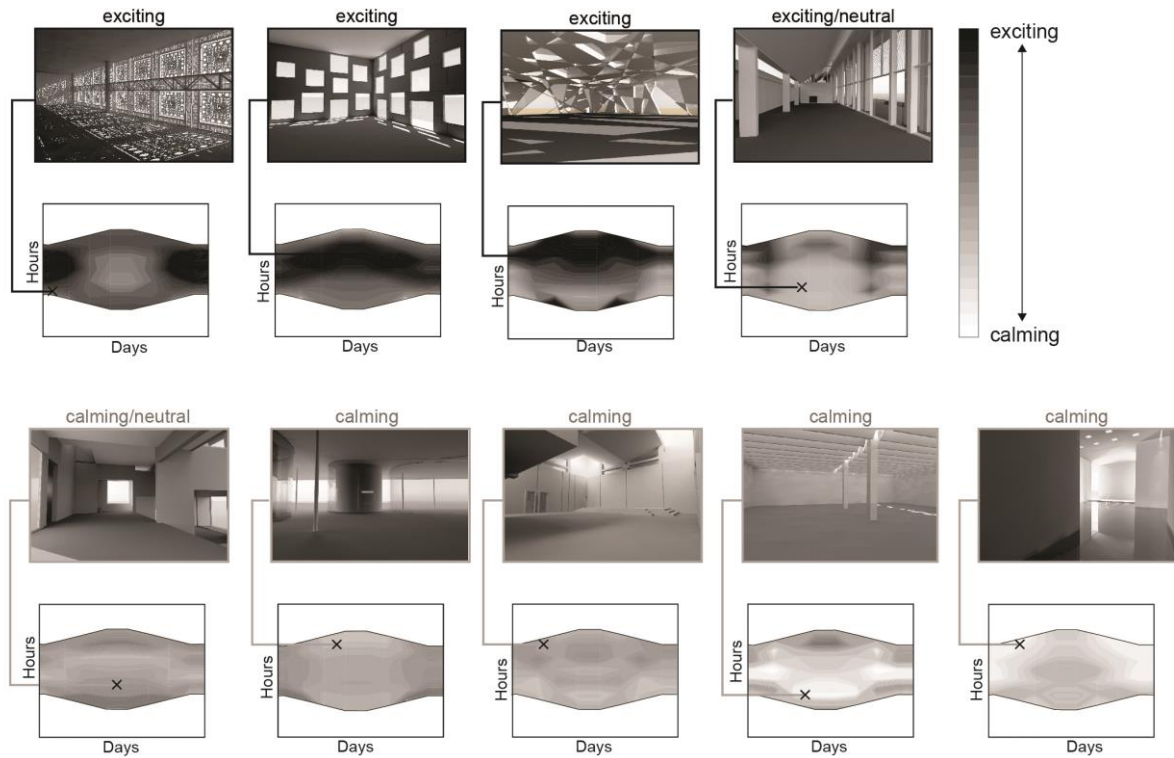


Figure 10. Application of the Modified Spatial Contrast measure applied to renderings of 56 symmetrical annual moments to predict ratings of calming (shown in light grey) or excitement (shown in dark grey) in all nine spaces used in the experiment.

4.5 Outlook

The authors acknowledge that this single view position, while commonly used in architectural practice to communicate occupant experience through interior renderings, is limited in its ability to capture the immersive reality of space. To further validate this approach, the authors will conduct a series of upcoming experiments with an expanded set of architectural spaces and view parameters. To

limit potential error due to screen size, brightness, and tone-mapping, these forthcoming experiments will be conducted under controlled laboratory conditions using screen technologies with an extended view (virtual reality) and luminance range. Existing studies in visual comfort have found a significant effect of luminance distribution on gaze patterns within the field-of-view⁴², which necessitates a more sophisticated approach to both the application and assessment of perceptual lighting effects. Further work is also needed to integrate knowledge of the visual system as our finding that some pixel resolutions are better predictors of perceptual effects than others suggests a possible psychophysical connection.

5. Conclusion

Building upon the authors' preliminary simulation-based studies which attempted to measure the compositional and temporal dynamics of daylight, this paper introduces the first data set which confirms a link between quantitative contrast measures, human perceptions of daylight composition in digital renderings and their varied effects over time. This experiment resulted in the following findings: 1) both space and sky condition have a significant effect on subject responses for ratings of contrast, direction, complexity, excitement, and stimulation, 2) local neighbourhood contrast measures such as RAMMG and specific levels within that metric (RAMM5, i.e. Modified Spatial Contrast) were found to be good predictors of contrast-based visual effects, especially ratings of diffuse – direct, calming – exciting and subdued – stimulating. Using the ordered logit model introduced in Section 4.4, this paper introduces a model for predicting the proportional odds of subject responses to impressions of calming – exciting. Forthcoming experiments will seek to validate whether this model is generalizable to a broader range of architectural spaces and immersive view parameters.

This model could allow architects and building performance engineers to calculate subject impressions of visual interest throughout a single view across a series of annual instances. While at each instance, a single point-in-time quantitative analysis may be less useful to designers who can evaluate this performance qualitatively, Modified Spatial Contrast becomes useful in its ability to predict dynamic effects which may be unanticipated. This annual prediction for excitement can indeed be combined with dynamic illuminance metrics and glare risk models. By predicting how visually engaging a space may be (and how this changes over time), this research offers a new dimension in daylight performance assessment. Rather than be satisfied with the knowledge that a space achieves enough or too much daylight, this model evaluates human perceptual responses to daylight composition. Ultimately, the assessment of perceptual daylight effects will have to be considered alongside existing illumination and glare-based comfort metrics on a dynamic scale to reach a truly holistic evaluation of daylight performance in architecture.

Funding

Funding for this research was provided by the École Polytechnique Fédérale de Lausanne (EPFL) and through a grant awarded by the Velux Stiftung Foundation (project 936).

Acknowledgements

A special thank you to Giorgia Chinazzo for providing geometric modelling support and Jan Wienold for his expert advice in Radiance.

References

- [1] Holl S. *Color Light Time*, Zurich, Switzerland: Lars Muller Publishers, 2011.
- [2] Pallasmaa J. *The Eyes of the Skin: Architecture and the Senses*, 2nd edition. Hoboken NJ: Wiley and Sons, 2005.
- [3] Steane MA. *The Architecture of Light: Recent Approaches to Designing with Natural Light*. New York: Routledge, 2011.
- [4] Reinhart C, Mardaljevic JRZ. Dynamic daylight performance metrics for sustainable building design. *Leukos* 2006; 3(1): 7-31.
- [5] Konis K, Lee E, ClearR. Visual comfort analysis of innovative interior and exterior shading systems for commercial buildings using high resolution luminance images. *Leukos* 2011; 7(3): 167-188.
- [6] Wienold J, Christoffersen J. Evaluating methods and development of a new glare prediction model for 152 daylight environments with the use of CCD cameras. *Energy and Buildings* 2006; 38(7): 743-757.
- [7] Cuttle C. *Lighting Design: A Perception-based Approach*. New York: Routledge, 2015.
- [8] Flynn J, Hendrick C, Spencer T, Martyniuk O. A guide to methodology procedures for measuring subjective impressions in lighting. *Journal of Illuminating Engineering Society* 1979; 8: 95-110.
- [9] Wymelenberg K, Inanici M. The effect of luminance distribution patterns on occupant preference in a daylighted office environment. *Leukos* 2010; 7(2): 103 - 122.
- [10] Newsham G, Richardson C, Blanchet C, Veitch J. Lighting quality research using rendered images of offices. *Lighting Research and Technology* 2005; 37(2): 93-112.
- [11] Newsham G, Cetegen D, Veitch J, Whitehead L. Comparing lighting quality evaluations of real scenes with those from high dynamic range and conventional images. *ACM Transactions on Applied Perception* 2010; 7(2): 1-25.
- [12] Veitch J, Newsham G. Preferred luminous conditions in open plan offices: Research and practice recommendations. *Lighting Research and Technology* 2000; 32(4): 199-212.
- [13] Loe D, Mansfield K, Rowlands E. Appearance of lit environment and its relevance in lighting design: Experimental study. *Lighting Research and Technology* 1994; 26(3): 119-133.

- [14] Cetegen D, Veitch J, Newsham G. View size and office illuminance effects on employee satisfaction. In *Proceedings of Balkan Light*, Ljubljana, Slovenia, 2008: 243-252.
- [15] Boubekri M, Hull R, Boyce L. Impact of window size and sunlight penetration on office worker's mood and satisfaction: A novel way of assessing sunlight. *Environment and Behavior* 1991; 23(4): 474-493.
- [16] Tiller D, Veitch J. Perceived room brightness: Pilot study on the effect of luminance distribution. *Lighting Research and Technology* 1995; 27(2): 93-101.
- [17] Parpaire K, Baker N, Steemers K, Compagnon R. The luminance differences index: a new indicator of user preferences in daylight spaces. *Lighting Research and Technology* 2002; 34(1): 53-68.
- [18] Michelson A. *Studies in Optics*. Chicago: University of Chicago Press, 1927.
- [19] Pavel M, Sperling G, Riedl T, Vanderbeek A. Limits of visual communication: the effect of signal-to-noise ratio on the intelligibility of American Sign Language. *Journal of the Optical Society of America A* 1987; 4(12): 2355 - 2365.
- [20] Simone G, Pedersen M, Hardeberg JY. Measuring perceptual contrast in digital images. *Journal of Visual Communication and Image Representation* 2012; 23(3): 491-506.
- [21] Hess R, Bradley A, Piotrowski L. Contrast-coding in amblyopia. I. differences in the neural basis of human amblyopia. *Proceedings of Royal Society of London Series B* 1983; 217(1208): 309-330.
- [22] Tremeau A. Color contrast parameters for analysing image differences. In *Proceedings of the University of Derby Color Image Science 2000 Conference*, Derby, UK, 2000: 11-23.
- [23] Tadmor Y, Tolhurst D. Calculating the contrasts that retinal ganglion cells and LGN neurones encounter in natural scenes. *Vision Research* 2000; 40: 3145-3157.
- [24] Rizzi A, Algeri G, Medeghini D, Marini A. A proposal for contrast measure in digital images. In *Proceedings of Second European Conference on Color in Graphics, Imaging and Vision*, Aachen, Germany, 2004.
- [25] Matekovic K, Neumann L, Neumann A, Psik T, Purgathofer W. Global contrast factor: a new approach to image contrast. In *Proceedings of the First Eurographics conference on Computational Aesthetics in Graphics, Visualization and Imaging*, Aire-la-ville, Switzerland, 2005: 159-167.
- [26] Rockcastle S, Andersen M. Measuring the dynamics of contrast and daylight variability in architecture: A proof-of-concept methodology. *Building and Environment*, 2014; 81: 320-333.
- [27] Adelson E, Anderson C, Bergen J, Burt P, Ogden J. Pyramid methods in image processing. *RCA Engineer*, 1984; 29(6): 33-41.
- [28] Simone G, Pedersen M, Hardeberg JY, Rizzi A. Measuring perceptual contrast in a multi-level framework. In *Proceedings of Electronic Imaging 2009*, San Jose, California, 2009.
- [29] Rizzi A, Simone G, Cordone R. A modified algorithm for perceived contrast measure in digital images. In *Proceedings to Conference on Colour in Graphics, Imaging, and Vision 2008*, Barcelona, Spain, 2008: 249-252.
- [30] Newsham G, Marchand R, Veitch J. Preferred surface luminance in offices: A pilot study. In *Proceedings of the IESNA Annual Conference*, Salt Lake City, Utah, 2002: 375-398.

- [31] Cauwerts C, Bodart M. Investigation of 3D projection for qualitative evaluation of daylight spaces. In *Proceedings of PLEA 2011 - 27th, Conference of Passive and Low Energy Architecture*, Louvain-la-Neuve, Belgium, 2011.
- [32] Cauwerts C, Bodart M, Labayrade R. Assessing lighting appearance using pictures: Influence of tone-mapping parameters and lighting conditions in the visualization room. In *Proceedings of CISBAT*, Lausanne, Switzerland, 2013: 359-364.
- [33] Whitehead L, Ward G, Stuerzlinger W, Seetzen H. *High Dynamic Range Display Devices*. United States Patent 6,891,672, 2005.
- [34] Cauwerts C. *Influence of presentation modes on visual perceptions of daylight space*. Ph.D. dissertation, Faculty of Architecture, Architectural Engineering, and Urbainsm, Universite Catholique de Louvain, Louvain-la-Neuve, Belgium, 2013.
- [35] Villa C, Labayrade R. Validation of an online protocol for assessing the luminous environment. *Lighting Research and Technology*, 2013; 45(4): 401-420.
- [36] Ward G. The RADIANCE Lighting Simulation and Rendering System. In *Proceedings of '94 SIGGRAPH Conference*, Orlando, Florida, 1994: 459-472.
- [37] Rockcastle S, Andersen M. Celebrating contrast and daylight variability in contemporary architectural design: A typological approach. In *Proceedings of 12th European Lighting Conference LUX EUROPA 2013*, Krakow, Poland, 2013: 345-350.
- [38] Ward G, Rushmeier H, Piatko C. *A Visibility Matching Tone Reproduction Operator for High Dynamic Range Scenes*. Berkeley, CA: Lawrence Berkeley National Laboratory, 1997.
- [39] Rockcastle S, Andersen M. Human perceptions of daylight composition in architecture: A preliminary study to compare quantitative contrast measures with subjective user assessments in HDR renderings. In *Proceedings of 14th International Conference of the International Building Performance Simulation Association*, Hyderabad, India, 2015: 1205-1212.
- [40] Vogels I. Atmosphere metrics: Development of a tool to quantify experienced atmosphere. In J. Westerink, M. Ouwerkerk, T. Overbeek and W. Pasveer (eds) *Probing Experience: From Assessment of User Emotions and Behaviour to Development of Products*, Dordrecht, The Netherlands: Springer, 2008: pp. 25-41.
- [41] Demers C. A classification of daylighting qualities based on contrast and brightness analysis. In *Conference Proceedings of the American Solar Energy Society*, Cleveland, Ohio, 2007.
- [42] Sarey Khanie M, Stoll J, Einhauser W, Wienold J, Andersen M. Gaze-driven approach for estimating luminance values in the field of view for discomfort assessments. In *Proceedings of CIE 2015*, Manchester UK, 2015

Appendix

The centre and surround components of the Difference of Gaussian metric (DOG)²³ in Equation (9) are shown below. The center component $R_c(x, y)$ can be described as,

$$R_c(x, y) = \sum_{i=x-3r_c}^{x+3r_c} \sum_{j=y-3r_c}^{y+3r_c} G_{\text{Center}}(i-x, j-y) p_{i,j}, \quad (\text{i})$$

with a bi-dimensional Gaussian filter,

$$G_{\text{Center}}(x, y) = e^{-(x^2+y^2)/r_c^2}, \quad (\text{ii})$$

where $p_{i,j}$ are the pixel intensities at position (i, j) and r_c is the radius of the Gaussian filter $G_{\text{Center}}(x, y)$ with centre (x, y) .

The surround component $R_s(x, y)$ in Equation (9) can be described as a bi-dimensional Gaussian filter:

$$R_s(x, y) = \sum_{i=x-3r_s}^{x+3r_s} \sum_{j=y-3r_s}^{y+3r_s} G_{\text{Surround}}(i - x, j - y) p_{i,j}, \quad (\text{iii})$$

with a Gaussian filter with centre (x, y) and a radius r_s

$$G_{\text{Surround}}(x, y) = 0.85 \left(\frac{r_c}{r_s}\right)^2 e^{-(x^2+y^2)/r_s^2}, \quad (\text{iv})$$

where the $0.85 \left(\frac{r_c}{r_s}\right)^2$ sets the integrated sensitivity (or volume) of the surround component to be 85% of that of the centre.

Table 1. Pearson correlation coefficient ρ between the percentage of subjects who ranked the image 5-7 on each ordinal scale and existing contrast measures, as introduced in Section 2.2. Bold values are for $p < 0.0001$.

	Michelson	RMS	SC	RAMM5	RAMMG	DOG $r_c=1$ $r_s=2$	RSC $r_c=1$ $r_s=2$
contrast	0.28	0.50	0.36	0.69*	0.72*	0.37	0.11
uniformity	0.30	0.42	0.30	0.64	0.64	0.31	0.15
variation	0.44	0.36	0.34	0.58	0.60	0.19	0.19
direct	0.11	0.57	0.45	0.80*	0.74*	0.49	0.05
complex	0.30	0.58	0.60	0.72*	0.69*	0.40	0.11
exciting	0.17	0.62	0.50	0.79*	0.72*	0.42	0.12
stimulating	0.23	0.59	0.50	0.78*	0.71*	0.39	0.09

*Rating pair and contrast measurements with $\rho \geq 0.70$ ($p < 0.0001$) were considered most significant.